Emparejamiento de Ciclos y Estabilidad de Funciones de Rango: dos historias sobre Homología Persistente

Inés García Redondo

Seminario de Topological Data Analysis

Universidad Autónoma de Madrid

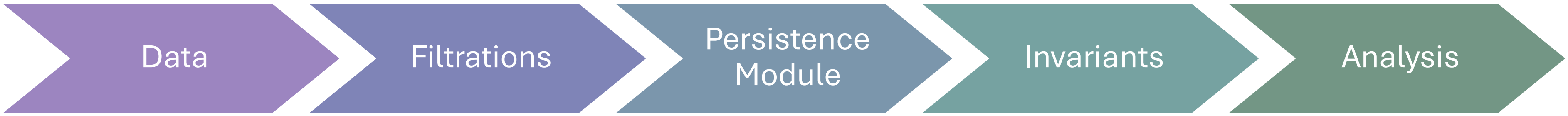
21/03/2024

IMPERIAL

EPSRC CENTRE FOR DOCTORAL TRAINING IN GEOMETRY AND NUMBER THEORY

LSGNT London School of Geometry & Number Theory

PH pipeline


Data
Filtrations
Persistence Module
Invariants
Analysis

PH pipeline

Data → Filtrations → Persistence Module → Invariants → Analysis

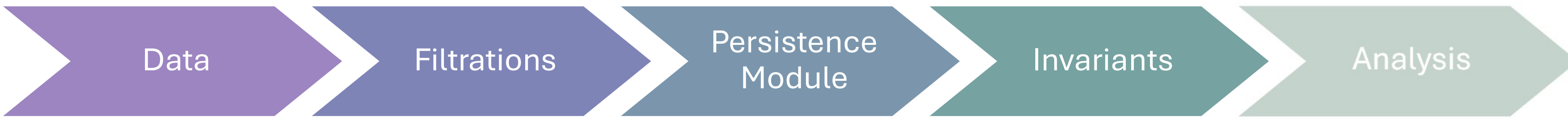
- **Filtration:** $F: (P, \leq) \to \mathrm{Simp},\ F(p) \subset F(q),\ p \leq q$
- **Persistence module:** $M: (P, \leq) \to \mathrm{Vec}_k$
  - Single-parameter: $(P, \leq) = (\mathbb{R}, \leq)$
  - Multiparameter: $(\mathrm{P}, \leq) = (\mathbb{R}^n, \preccurlyeq)$

Structure Theorem:
(Zomorodian and Carlsson, 2005; Crawley-Boevey, 2015)

$$M \simeq \mathrm{M}_1 \oplus \cdots \oplus \mathrm{M}_l$$

If $(P, \leq) = (\mathbb{R}, \leq) \Rightarrow M_j = \mathrm{I}[b_j, d_j)$

PH pipeline



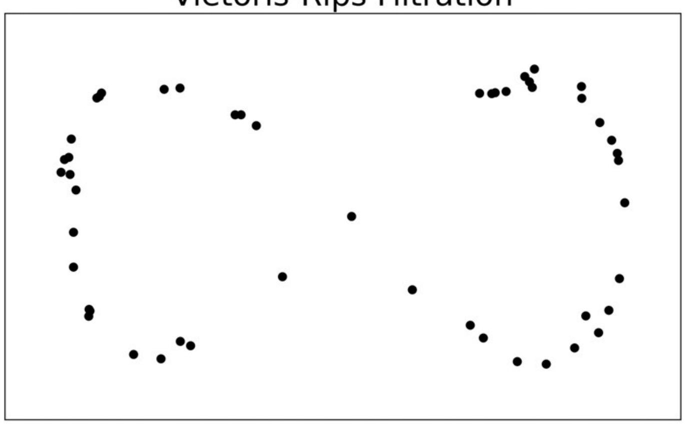
- **Filtration:** $F\colon (P, \leq) \to \mathrm{Simp},\ \ F(p) \subset F(q),\ \ p \leq q$
- **Persistence module:** $M\colon (P, \leq) \to \mathrm{Vec}_k$
  - Single-parameter: $(P, \leq) = (\mathbb{R}, \leq)$
  - Multiparameter: $(\mathrm{P}, \leq) = (\mathbb{R}^n, \preccurlyeq)$

Structure Theorem:
(Zomorodian and Carlsson, 2005; Crawley-Boevey, 2015)

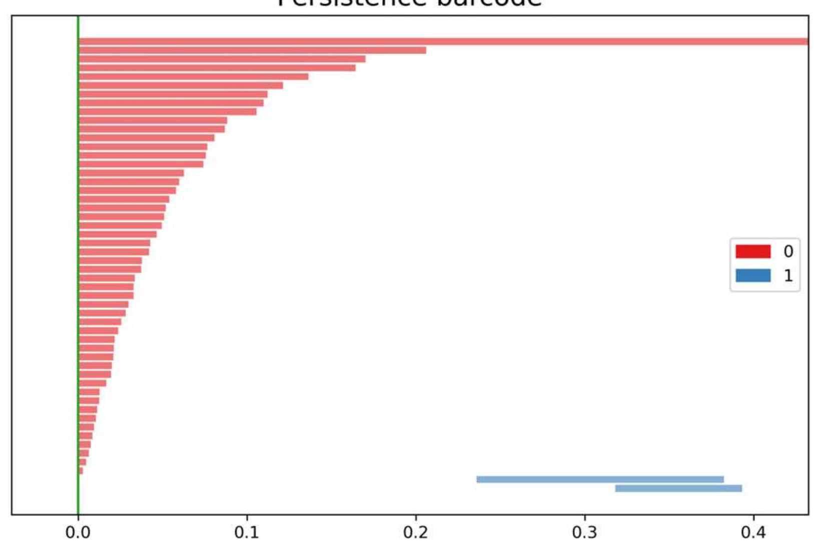
$$M \simeq \mathrm{M}_1 \oplus \cdots \oplus \mathrm{M}_l$$

If $(P, \leq) = (\mathbb{R}, \leq) \Rightarrow M_j = \mathrm{I}[b_j, d_j)$

Vietoris-Rips Filtration



Persistence barcode



PH pipeline

Data → Filtrations → Persistence Module → Invariants → Analysis

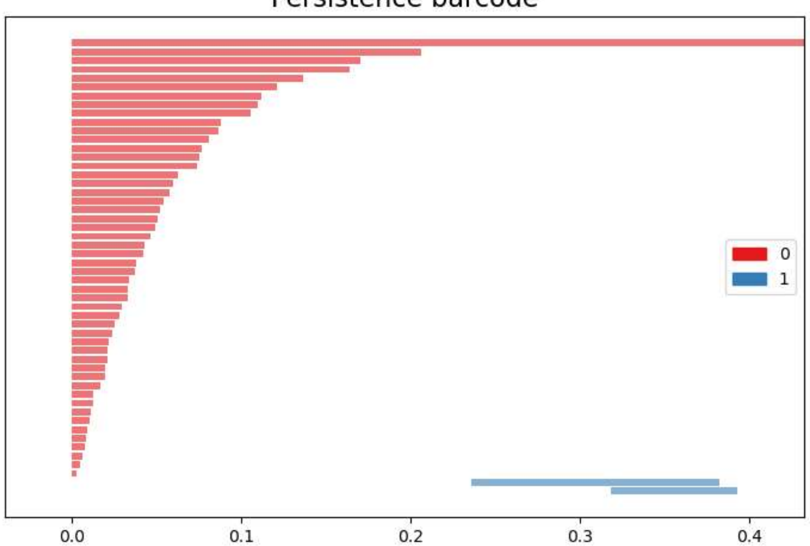
- **Filtration:** $F\colon (P, \leq) \to \mathrm{Simp}$
- **Persistence module:** $M\colon (P, \leq) \to \mathrm{Vec}$
  - Single-parameter: $(P, \leq) = (\mathbb{R}, \leq)$
  - Multiparameter: $(\mathrm{P}, \leq) = (\mathbb{R}^n, \preccurlyeq)$

Structure Theorem:
(Zomorodian and Carlsson, 2005; Crawley-Boevey, 2015)

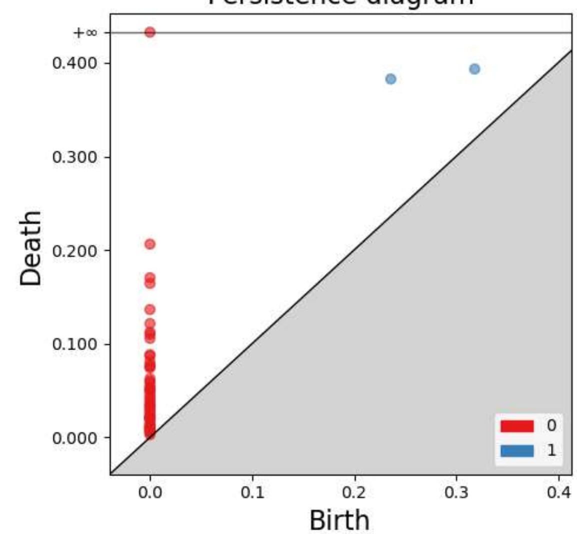
$$M \simeq \mathrm{M}_1 \oplus \cdots \oplus \mathrm{M}_l$$

If $(P, \leq) = (\mathbb{R}, \leq) \Rightarrow M_j = \mathrm{I}[b_j, d_j)$

Persistence barcode



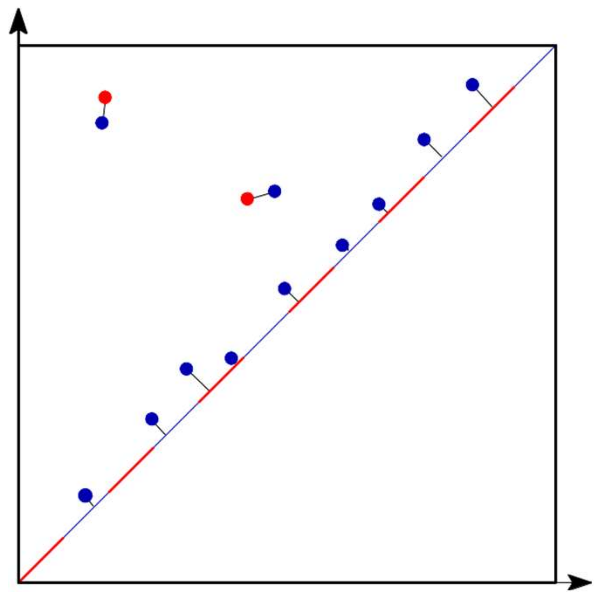
Persistence diagram



PH pipeline

Data → Filtrations → Persistence Module → Invariants → Analysis

Matching between PDs, from GUDHI documentation



Metrics over barcodes and PDs

- **Bottleneck distance:**

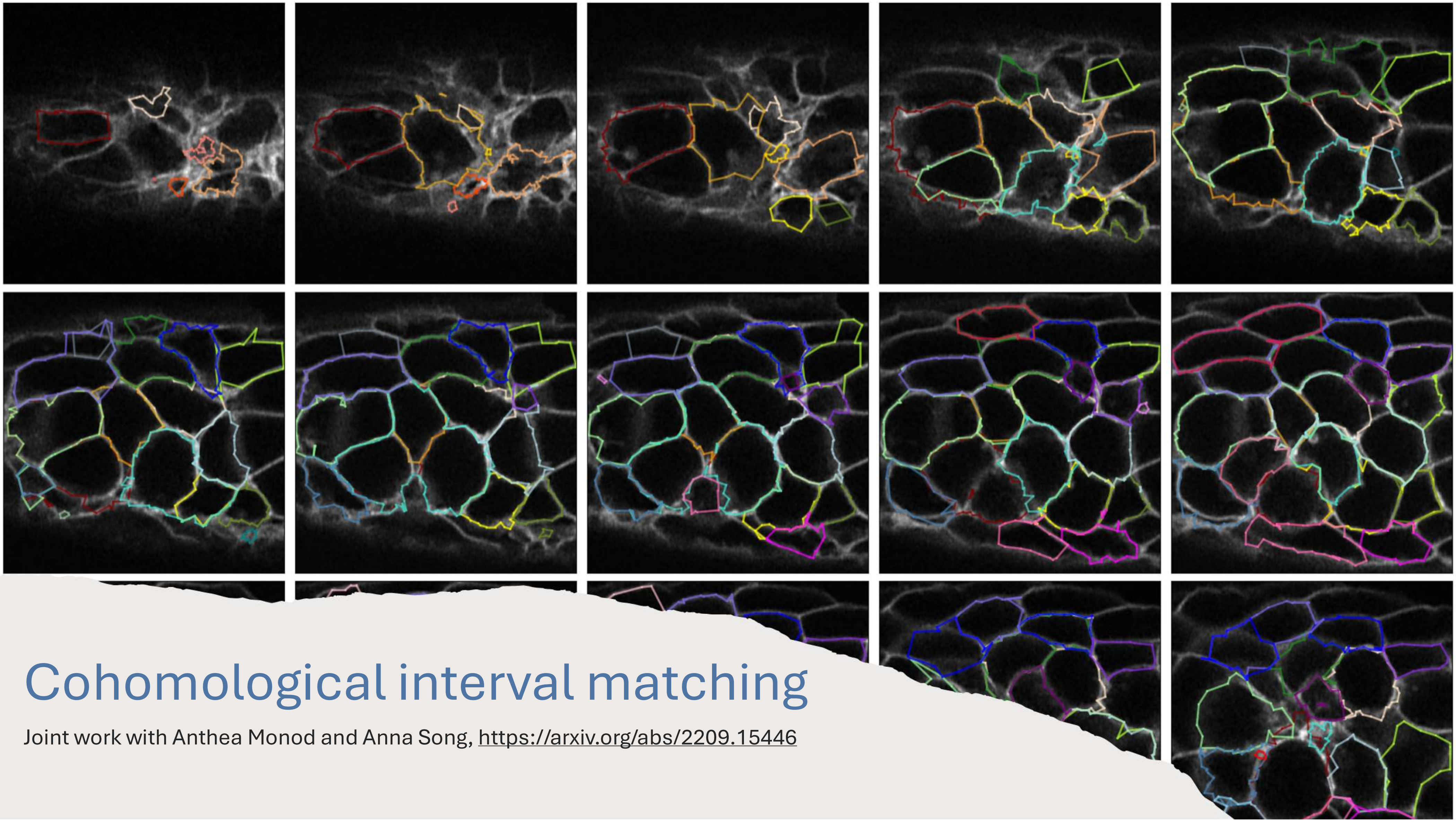
$$d_B(D_1, D_2) := \inf_{\phi: D_1 \to D_2} \sup_{x \in D_1} \|x - \phi(x)\|_\infty$$

- **Wasserstein distance:**

$$W_{p,q}(D_1, D_2) := \inf_{\phi: D_1 \to D_2} \left[\sum_{x \in D_1} \|x - \phi(x)\|_q^p \right]^{1/p}$$

Stability results:

$\textit{metric on invariants} \leq \textit{metric on input data}$

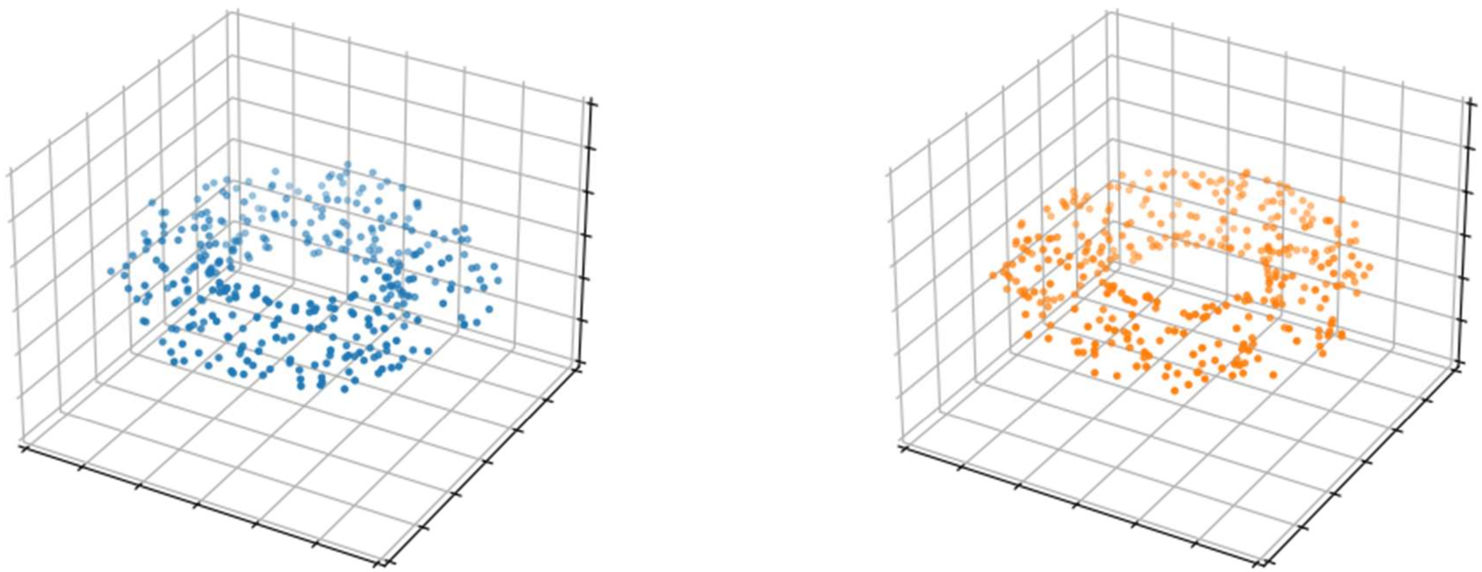


Cohomological interval matching

Joint work with Anthea Monod and Anna Song, https://arxiv.org/abs/2209.15446

Cycle Matching

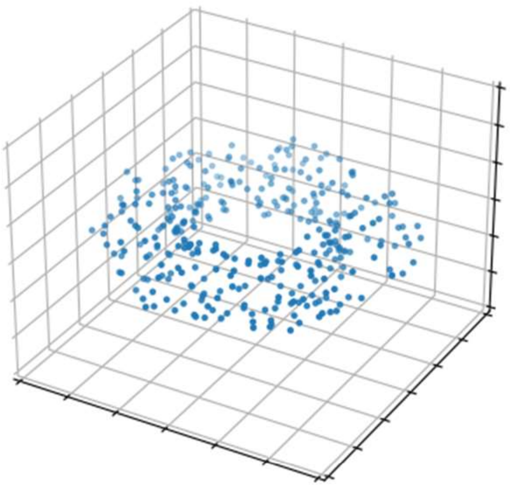
(Reani & Bobrowski, 2021)

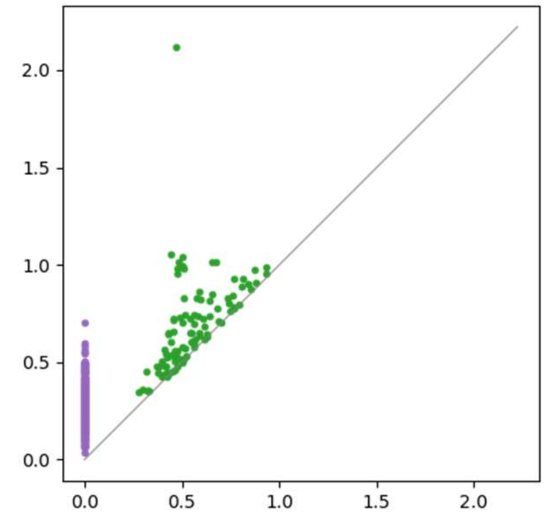


Two samples of N = 350 points with a gaussian noise of scale 0.1 added over a flat torus of radii R = 2 and r = 1

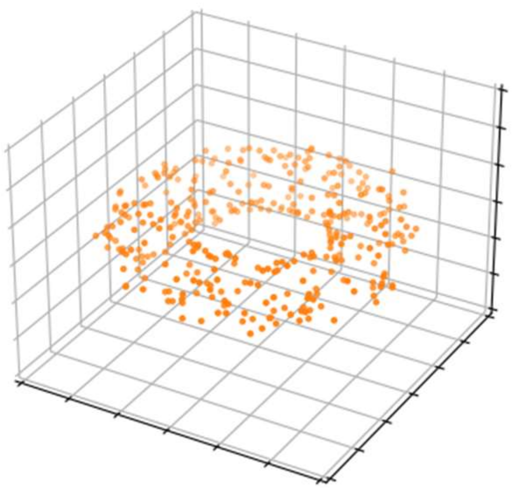
Cycle Matching

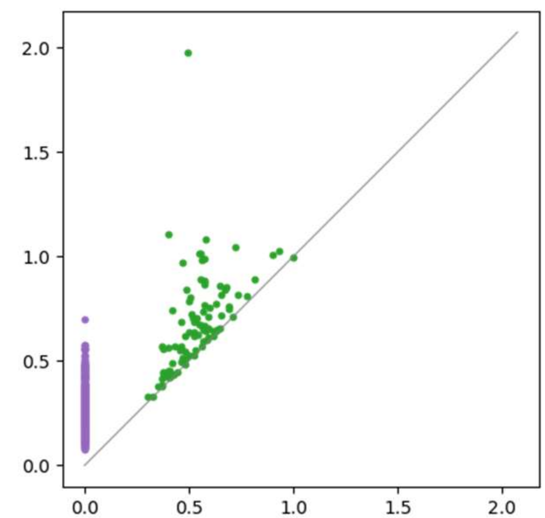
(Reani & Bobrowski, 2021)





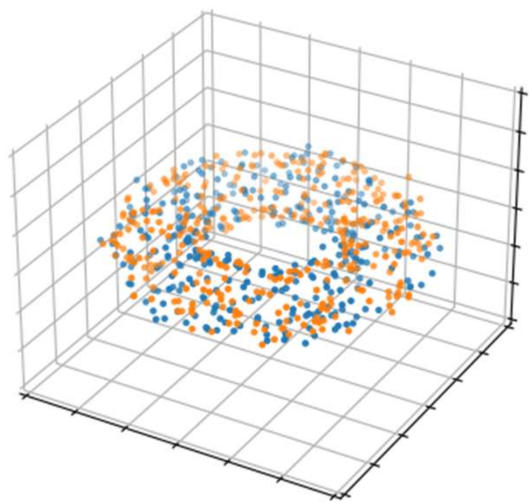
Obtain filtrations
+
Compute PH

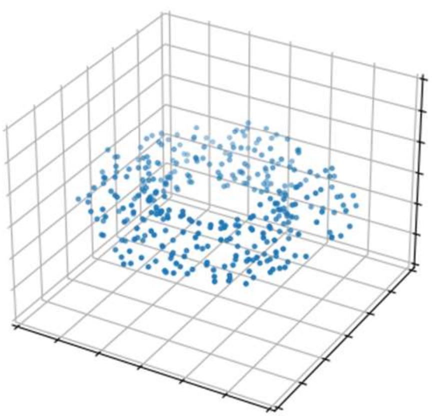




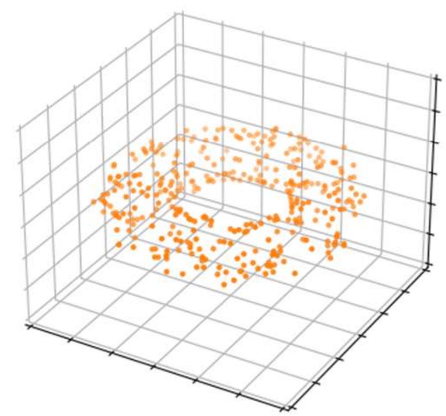
Cycle Matching

(Reani & Bobrowski, 2021)



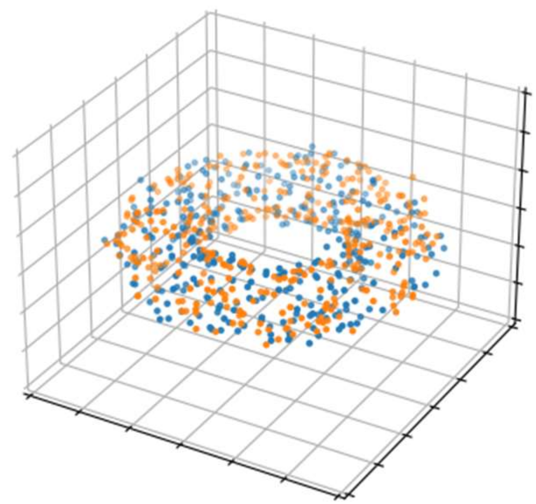


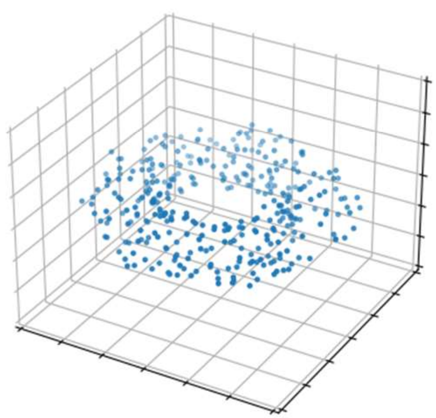
$$\begin{tikzcd} & Z = X \cup Y & \\ X \arrow[ur, "f"] & & Y \arrow[ul, "g"'] \end{tikzcd}$$



Cycle Matching

(Reani & Bobrowski, 2021)





$$
\begin{tikzcd}
 & Z_t \coloneqq VR_t(Z) & \\
X_t \coloneqq VR_t(X) \arrow[ur, "f_t"] & & Y_t \coloneqq VR_t(Y) \arrow[ul, "g_t"']
\end{tikzcd}
$$

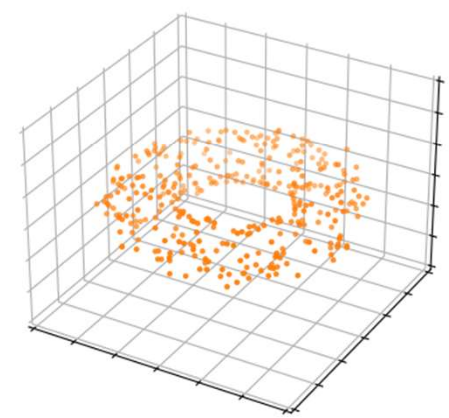


Image-persistence

f commutes with boundary maps

$$f_t \circ \partial_k^{X_t} = \partial_k^{Z_t} \circ f_t$$

(Cohen-Steiner et al., 2009)

$$\begin{tikzcd} Z_s \arrow[r, "i(s \leq t)"] & Z_t \\ X_s \arrow[u, "f_s"] \arrow[r] & X_t \arrow[u, "f_t"'] \end{tikzcd} \implies \begin{tikzcd} H_k(Z_s) \arrow[r, "i_*(s \leq t)"] & H_k(Z_t) \\ H_k(X_s) \arrow[u, "f_{s,*}"] \arrow[r] & H_k(X_t) \arrow[u, "f_{t,*}"'] \end{tikzcd}$$

Image-persistence module:
$f_*: H_k(X) \to H_k(Z)$
$i_*(s \leq t): \mathrm{Im}(f_{s,*}) \longrightarrow \mathrm{Im}(f_{t,*})$

$$\mathrm{Im}\, f_*(H_k(X)) \cong \frac{f\left(\mathrm{Ker}\, \partial_k^X\right)}{\mathrm{Im}(\partial_{k+1}^Z) \cap f(\mathrm{Ker}\, \partial_k^X)}$$

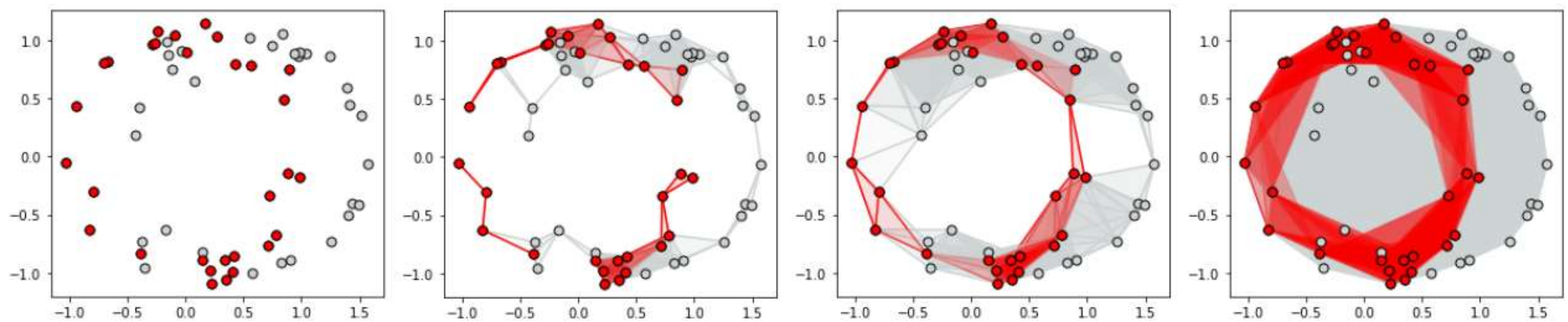


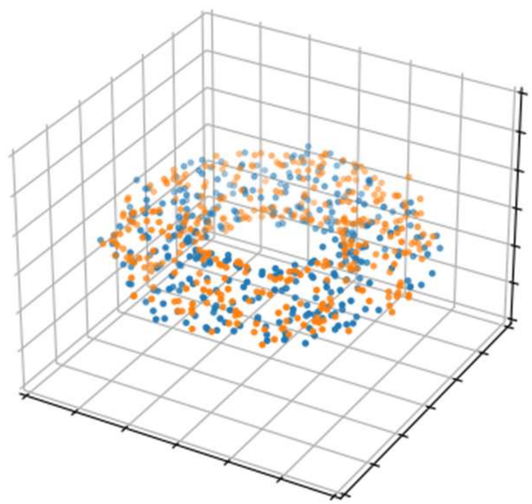
image of X in red | image-component killed | image-cycle created | image-cycle killed

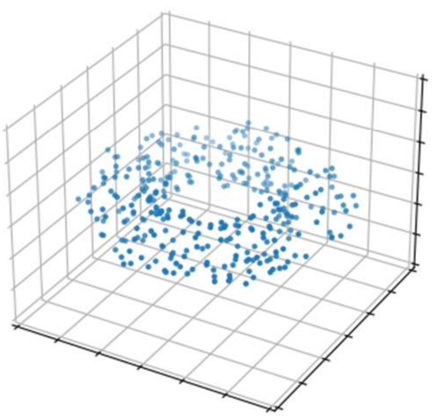
Here the 0 homology of X has 3 components but inside of Z its all one component

Here the 1 homology of X has 1 cycle but inside of Z this cycle is a boundary

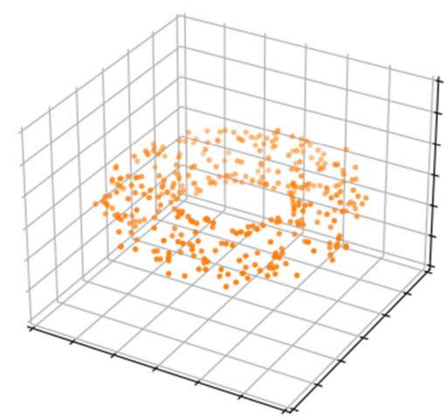
Cycle Matching

(Reani & Bobrowski, 2021)





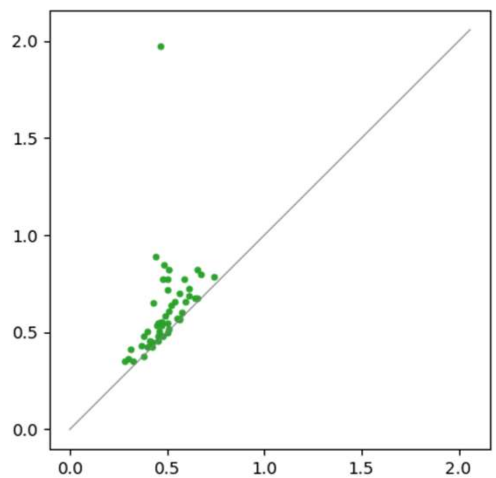
$$\begin{tikzcd} & H_*(Z_t) & \\ H_*(X_t) \arrow[ur, "(f_t)_*"] & & H_*(Y_t) \arrow[ul, "(g_t)_*"'] \end{tikzcd}$$

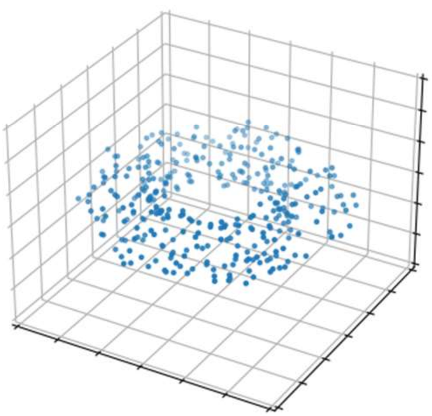


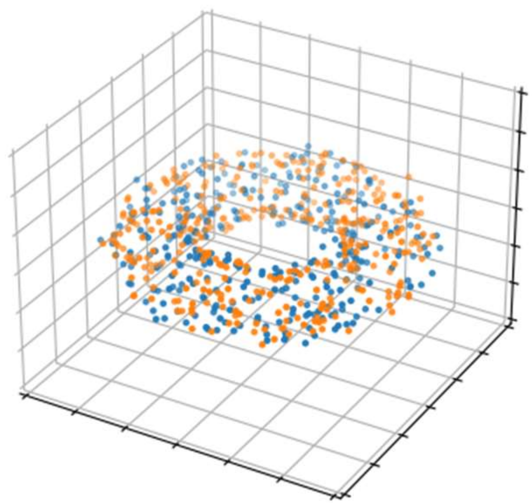
Cycle Matching

(Reani & Bobrowski, 2021)

Image-persistence of X inside of Z





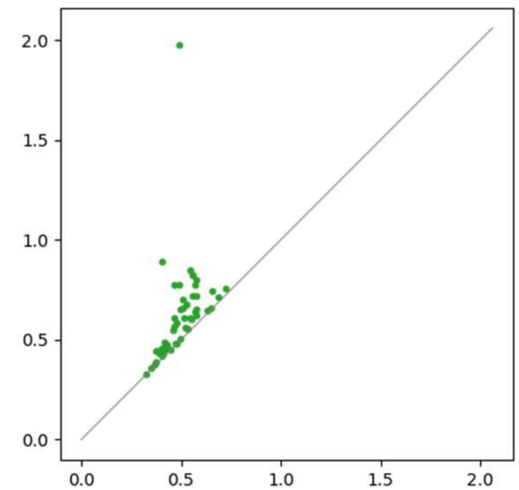


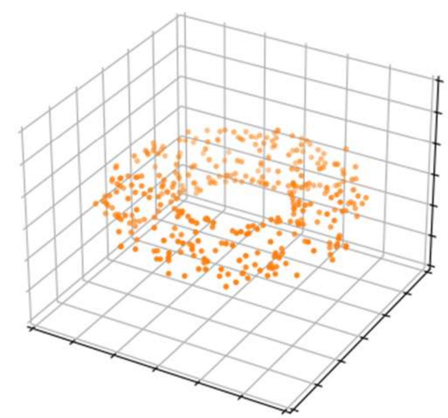
$$
\begin{tikzcd}
Im(f_t)_* \arrow[r, phantom, "\subset"] & H_*(Z_t) \arrow[r, phantom, "\supset"] & Im(g_t)_* \\
H_*(X_t) \arrow[u, "(f_t)_*"] & & H_*(Y_t) \arrow[u, "(g_t)_*"']
\end{tikzcd}
$$

Image-persistence

(cycles in X and Y up to boundaries in Z)

Image-persistence of Y inside of Z



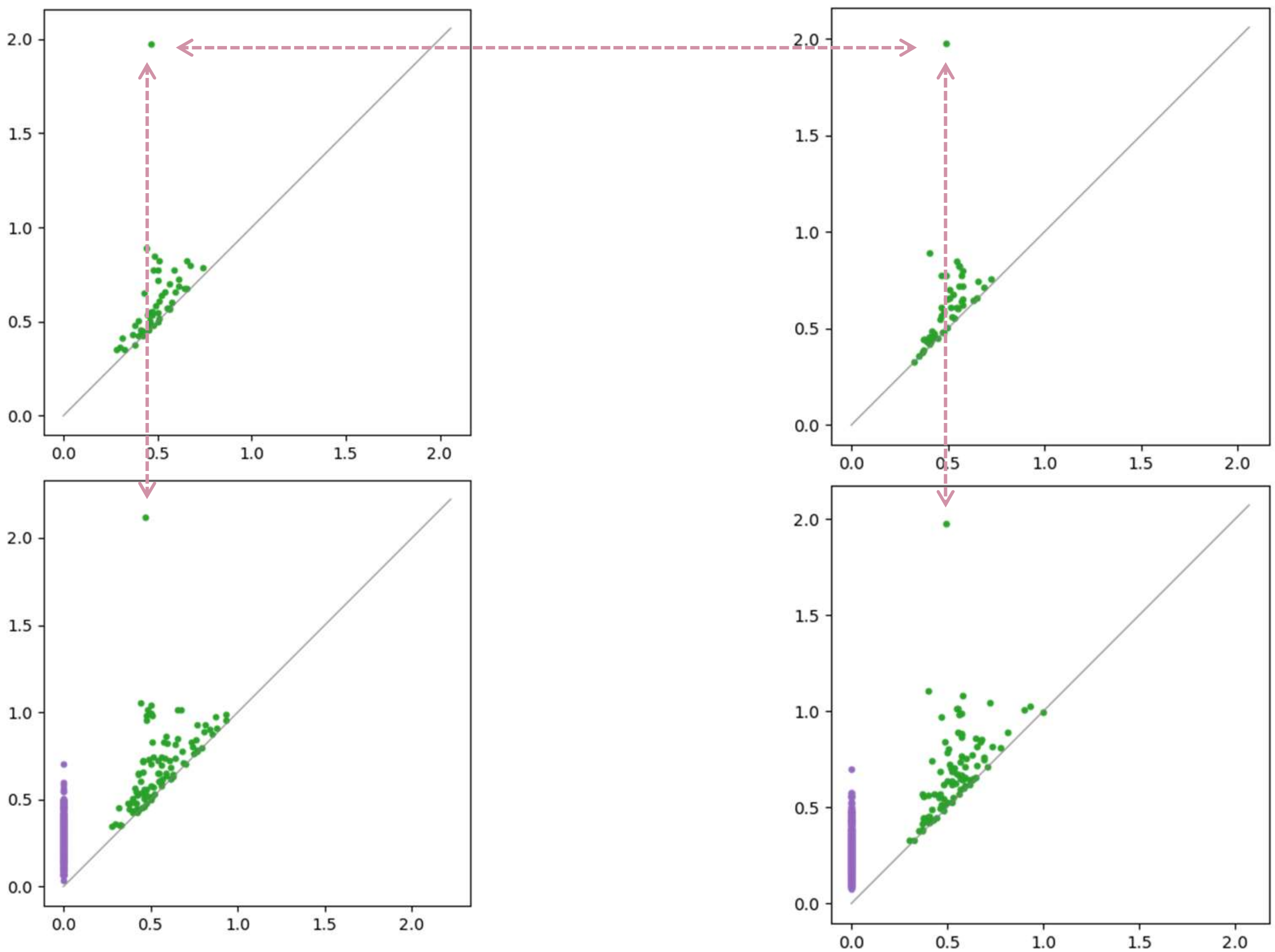


Cycle Matching

(Reani & Bobrowski, 2021)

Image-persistence of X inside Z

Image-persistence of Y inside Z



Persistence diagram of X

Persistence diagram of Y

PH computations

PH is expensive to compute

- Most naïve algorithm based on a matrix reduction algorithm via Gauss elimination
- Cubic complexity in number of simplices

Ripser (Bauer, 2021)

- State-of-the-art for PH computations
- Clearing algorithm
- Persistent cohomology
- Several other optimizations

Image-persistence

- Optimizations of ripser were extended
- Ripser-image (Bauer and Schmahl, 2022)

Fast cohomological interval matching

Joint work with Anthea Monod and Anna Song, https://arxiv.org/abs/2209.15446

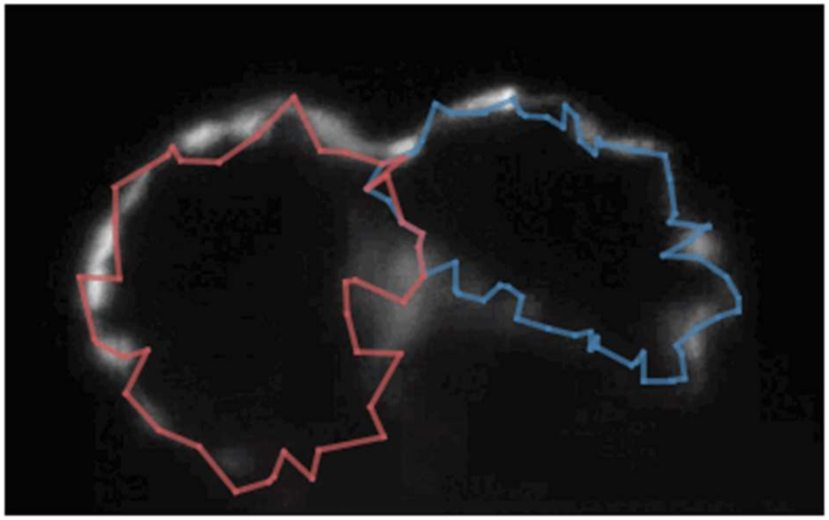
Theoretical

- Extended definitions of interval matching to make it compatible with a wider spectrum of filtrations

Practical

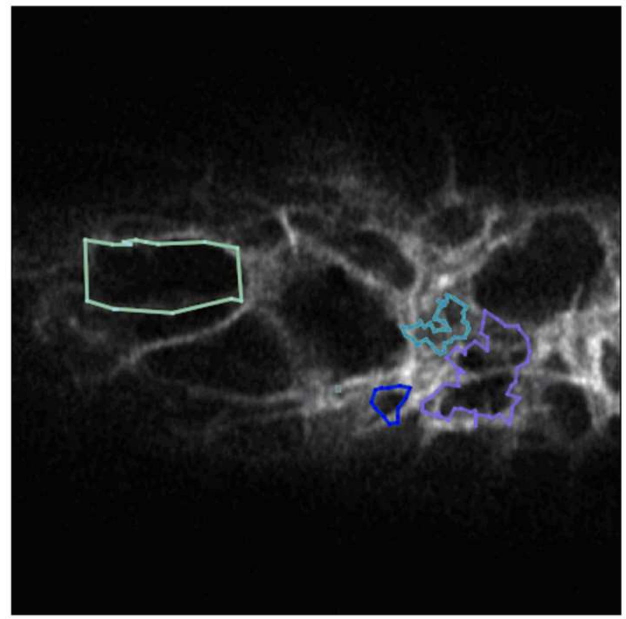
- Provided state-of-the-art-code for interval matching
- Proposed a parallelisation pipeline that accelerated computations to minutes using the HPC facilities
- Showcased real-life applications o this machinery

Tracking topological features

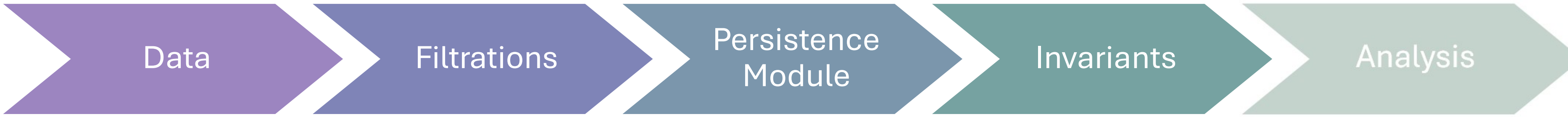


Video of primitive heartbeat from an embryo zebrafish, from Scherz et al. (2008)

Stack of 2D images of the primitive Lateral Line from an embryo zebrafish, from Hartmann et al. (2008)



PH pipeline



Let $\mathbb{R}^{2+} := \{(x,y) \in \{-\infty\} \cup \mathbb{R} \times \mathbb{R} \cup \{\infty\} : x \leq y\}$.

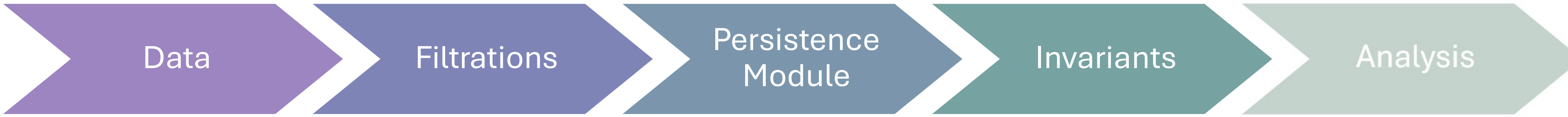
Definition

Given $M : \mathbb{R} \to \mathrm{Vec}$ p.f.d. persistence module, its *rank function* is defined as

$$\begin{array}{rccl} \beta^M : & \mathbb{R}^{2+} & \to & \mathbb{Z} \\ & (x,y) & \mapsto & \operatorname{rank} M(x \leq y) = \dim \operatorname{Im}\left(M(x \leq y)\right). \end{array}$$

- Introduced in the early 1990s in the work of Frosini as "**size functions**"
- Landi and Frosini (1997) provided an algebraic reinterpretation of size functions in terms of formal series that allowed them to construct several pseudo-distances on the size function space
  - **Deformation pseudo-distance**
  - **Haussdorff pseudo-distance** (renamed by d'Amico et al. (2003, 2006, 2010) as the matching distance)
  - L^p **pseudo-distance**
- Later this was reinterpreted in persistence theory: **barcodes prevailed because of rigorous metrics**

PH pipeline



Let $\mathbb{R}^{2+} := \{(x, y) \in \{-\infty\} \cup \mathbb{R} \times \mathbb{R} \cup \{\infty\} : x \preceq y\}$.

Definition

Given $M : \mathbb{R} \to \mathrm{Vec}$ p.f.d. persistence module, its *rank function* is defined as

$$\begin{array}{rccl} \beta^M : & \mathbb{R}^{2+} & \to & \mathbb{Z} \\ & (x, y) & \mapsto & \mathrm{rank}\, M(x \leq y) = \dim \mathrm{Im}\,(M(x \leq y)). \end{array}$$

- $T = \{t_1, \dots, t_\ell\} \subset \mathbb{R}$ discrete set of values over which the persistence module changes;
- $\overline{T} = T \cup \{-\infty, +\infty\}$;
- $S = \{s_0, s_1, \dots, s_\ell\}$ set of real numbers such that $s_{i-1} \leq t_i \leq s_i$;
- $s_{-1} = t_0 = -\infty$ and $s_{\ell+1} = t_{\ell+1} = +\infty$

$$\mu_i^j := \beta(s_{i-1}, s_j) - \beta(s_i, s_j) + \beta(s_i, s_{j-1}) - \beta(s_{i-1}, s_{j-1}). \tag{1}$$

PH pipeline

Data → Filtrations → Persistence Module → Invariants → Analysis

Let $\mathbb{R}^{2+} := \{(x, y) \in \{-\infty\} \cup \mathbb{R} \times \mathbb{R} \cup \{\infty\} : x \preceq y\}$.

Definition

Given $M : \mathbb{R} \to \mathrm{Vec}$ p.f.d. persistence module, its *rank function* is defined as

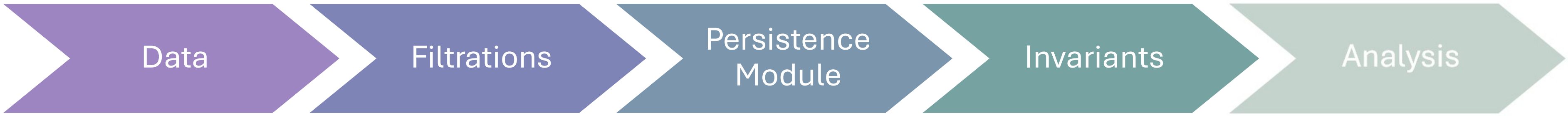
$$\begin{array}{rccl} \beta^M : & \mathbb{R}^{2+} & \to & \mathbb{Z} \\ & (x, y) & \mapsto & \operatorname{rank} M(x \leq y) = \dim \operatorname{Im}(M(x \leq y)). \end{array}$$

Definition (Persistence Diagram)

$$\operatorname{Dgm}(M) := \{(t_i, t_j) \in \overline{T} \times \overline{T} : t_i < t_j\} \cup \partial$$

- each point (t_i, t_j) has multiplicity μ_i^j;
- all the points in the diagonal $\partial = \{(x, y) \in \mathbb{R}^{2+} : x = y\}$ counted with infinite multiplicity.

PH pipeline



What about multiparameter persistence?

Structure Theorem:
(Zomorodian and Carlsson, 2005; Crawley-Boevey, 2015)

$$M \simeq \mathrm{M}_1 \oplus \cdots \oplus \mathrm{M}_l$$

- If $(P, \leq) = (\mathbb{R}, \leq) \Rightarrow M_j = \mathrm{I}[b_j, d_j)$
- If $(P, \leq) = (\mathbb{R}^n, \preccurlyeq)$, indecomposables have wild representation type

No hope for a direct, analogous definition of barcode on MPH (Carlsson & Zomorodian, 2009)

PH pipeline

Data → Filtrations → Persistence Module → Invariants → Analysis

What about multiparameter persistence?

Let $\mathbb{R}^{2n+} := \{(x, y) \in (\{-\infty\} \cup \mathbb{R})^n \times (\mathbb{R} \cup \{\infty\})^n : x \preceq y\}$.

Definition

Given $M : \mathbb{R}^n \to \text{Vec}$ p.f.d. persistence module, its *rank invariant* is defined as

$$\begin{array}{rccl} \beta^M : & \mathbb{R}^{2n+} & \to & \mathbb{Z} \\ & (x, y) & \mapsto & \operatorname{rank} M(x \preceq y) = \dim \operatorname{Im} \left(M(x \preceq y)\right). \end{array}$$

The space of rank invariants for n-dimensional persistence modules will be denoted by $\mathcal{I}_n$.

Generalised persistence diagrams

Patel (2018), Kim and Mémoli (2021), McCleary and Patel (2022)

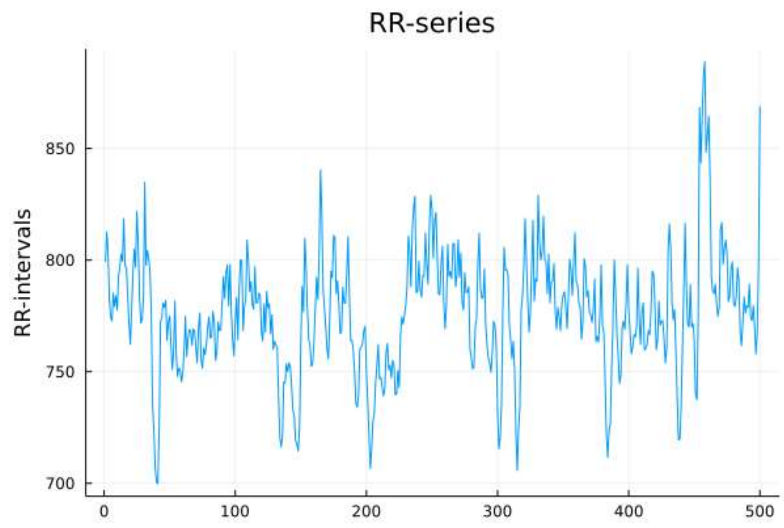
Rank functions and rank invariants in Data Analysis

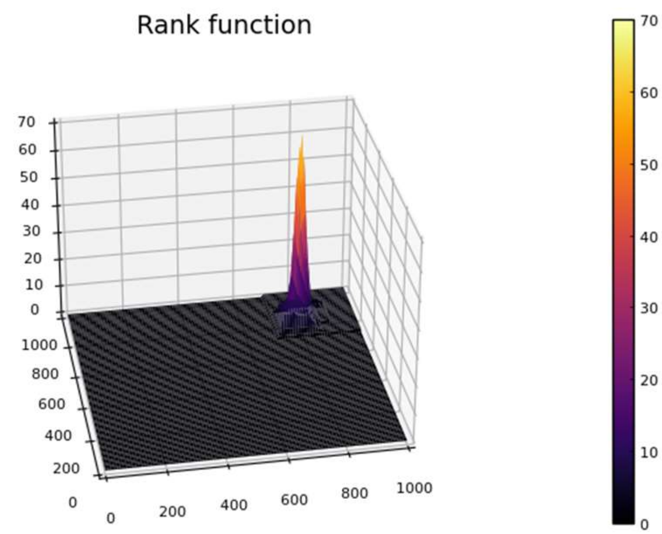
Advantages

- Easily generalisable to MPH.
- Computable from PDs and using RIVET (https://github.com/Qiquan-Wang/rank_stability)
- Benefitting from the methods of **Functional Data Analysis** (Ramsay 2002, 2005).
    - Tools to avoid the curse of dimensionality

Why rank invariants are not as popular as barcodes in Data Analysis?

1. **Theoretical issues:** less established stability theory
2. **Practical issues:** not so well known how they perform in data analysis task
    - Implemented in a descriptive statistics setting for FPCA by Robins and Turner (2016)
    - What about inferential, supervised learning tasks?





Computable Stability for Persistence Rank Functions ML

Joint work with Qiquan Wang, Pierre Faugère, Anthea Monod and Gregory Henselman-Petrusek, https://arxiv.org/abs/2307.02904

Functional Data Analysis using Rank Functions

Rank function based FSVM

- Distinguish healthy individuals from post-stroke patients
- **Input data:** RR-sequences
- Performance $> 80\%$, better than non-persistence-based methods and on par with more elaborated persistence-based techniques (Graff et al., 2021)

Hypothesis testing for biparameter rank functions

- Assess the impact of various types of noise on data
- **Input data:** point clouds with different noise added
- Validate the resilience to outliers of biparameter rank functions compared with single-parameter rank functions

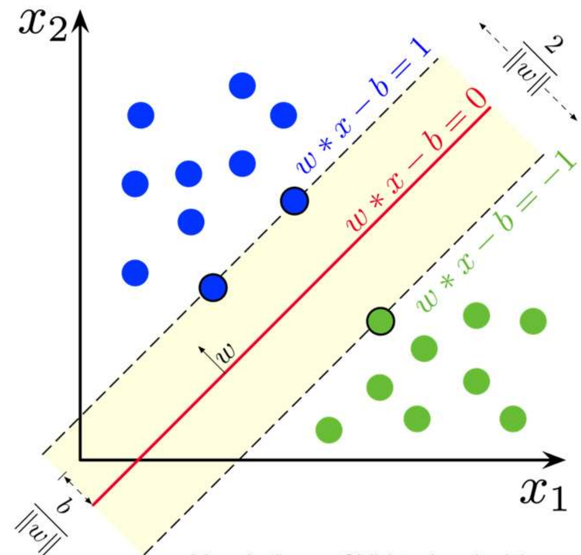
Lung tumour classification for biparameter rank functions

- Implemented non-parametric supervised methods (k-NN and Functional Maximum depth) for lung tumour classification
- **Input data:** Computed Tomography images
- Improved results with respect to previous persistence-based methods (Vandaele et al., 2023)

Method 1: FSVM

Classical SVM (Boser et al., 1992)

- Data: $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, $y_i = \pm 1$ depending on the class $\mathbf{x}_i \in \mathbb{R}^p$ belongs to.
- Hyperplane: $\mathbf{w}^\intercal \mathbf{x} + b = 0$, $\mathbf{w} \in \mathbb{R}^p$ normal vector.
- **Hard-margin:** if the data is *linearly separable* we can find two hyperplanes separating the two classes that maximise the area between them. This area is the *margin*.
- **Soft margin:** we allow some leeway for wrong classification.
- **Kernel trick** for non-linear boundaries: replace dot product by kernel functions. Popular kernels include polynomial and Gaussian Radial Basis Function (GRBF) kernels.



Margin for an SVM trained with samples from two classes, from Wikipedia

FSVM (Rossi et al., 2005)

- Data: $(f_1, y_1), \ldots, (f_n, y_n)$, with $f_i \in \mathcal{H}$ a Hilbert space and $y_i = \pm 1$ depending on the class of f_i.
- Same formulations as above.
- Kernels can also be extended to this setting.

FSVM on Single-Parameter Rank Functions

Data

- 86 sequences of 512 RR series (beat-to-beat time intervals) extracted from ECGs
- 2 groups of people in a similar age category:
  - 46 healthy individuals
  - 40 patients with recent stroke episodes

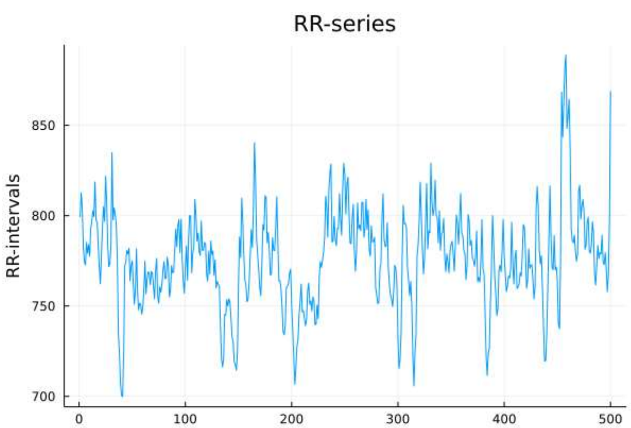
Topological representation

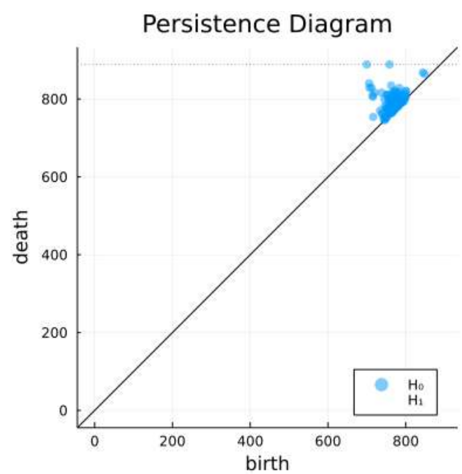
- Sublevel set filtrations from the RR series based on height function of y-coordinate.
- Computed 0-dimensional PH and obtained $\beta_1, \ldots, \beta_{86}$ rank functions
- Centred the functions to have mean 0

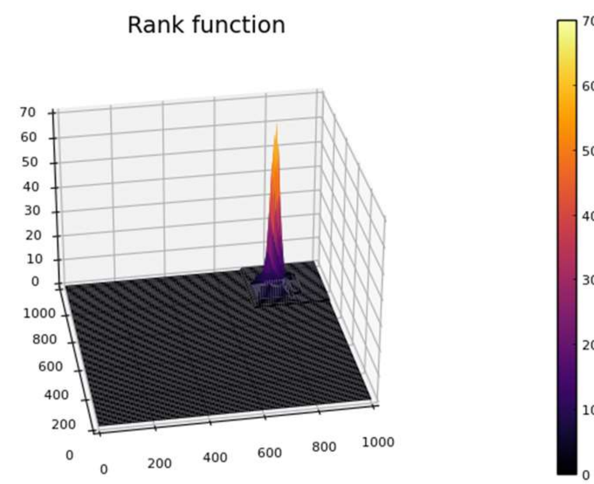
FSVM

Trained FSVM (soft-margin approach) classifiers using
- Linear kernel
- GRBF kernel
- Polynomial kernel $d = 2, 3, 5$

Input
- Original rank functions
- FPCA
- Haar Wavelet basis







FSVM on Single-Parameter Rank Functions

Kernel	Discretized rank functions		Projected rank functions on			
			FPCA basis functions		Wavelet basis functions	
	Accuracy	AUC-ROC	Accuracy	AUC-ROC	Accuracy	AUC-ROC
Linear	39.50%	0.408	66.80%	0.681	39.70%	0.404
GRBF	75.80%	0.762	50.30%	0.502	75.00%	0.756
Polynomial (d=2)	82.60%	0.829	84.20%	0.842	84.00%	0.842
Polynomial (d=3)	81.60%	0.829	82.00%	0.829	80.30%	0.816
Polynomial (d=5)	76.00%	0.775	75.50%	0.774	76.60%	0.786

Accuracy and AUC-ROCs of classifiers over ten iterations of five-fold cross validation.

- **Non-persistence approaches** using heart rate variability frequency and time domain parameters (Graff et al., 2021) only achieved AUC-ROCs of **0.79 and 0.75** respectively
- A **persistence-based approach**, based on selecting a wide range of topological indices based on vectorisations of the PD (Graff et al., 2021) has AUC-ROC **0.83**, on par with our method.

Method 3: FDA supervised classification methods

k-Nearest Neighbors (Cover and Hart, 1967)	Functional Maximum Depth (López-Pintado and Romo, 2009)
• Classification technique in both multivariate and functional data • The class of a new point is based on majority vote of k-closest neighbors. • This method is adaptable to different metric spaces: we work with rank invariants and the L^2 metric.	• For a collection of functions $f_1 \ \ldots, f_n : X \to \mathbb{R}$, define a *band* $B(f_1, \ldots, f_n) \coloneqq \left\{(x, y) : \ x \in X, \min_{i=1,\ldots,n} f_i(x) \leq y \leq \max_{i=1,\ldots,n} f_i(x)\right\}$ • *Band depth:* number of times a function lies in the band formed by a subcollection of functions. • Given classes of functions and a new function f, f will be assigned to the class that maximizes the band depth.

Lung tumour classification using rank invariants

Data

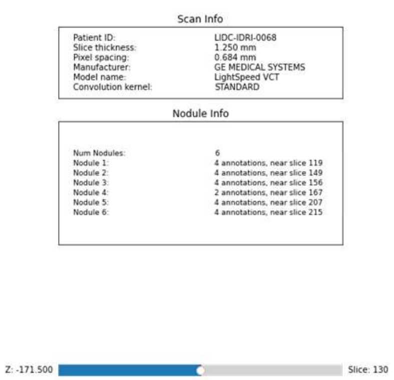
- 115 chest CT scans
  - 29 benign cases (24 with contrast)
  - 41 primary malignant (17 with contrast)
- Following Vandaele et al. (2023) converted scans to point clouds of landmarks on the tumour surfaces

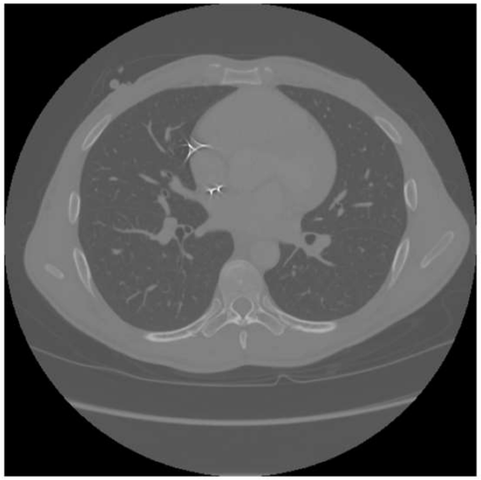
Topological representations

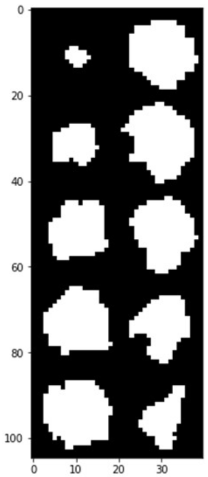
- Degree-Rips filtration
- Vietoris-Rips combined with the x, y and z height functions, building up the tumour in those directions
- Computed bi-parameter rank invariants (0-homology)

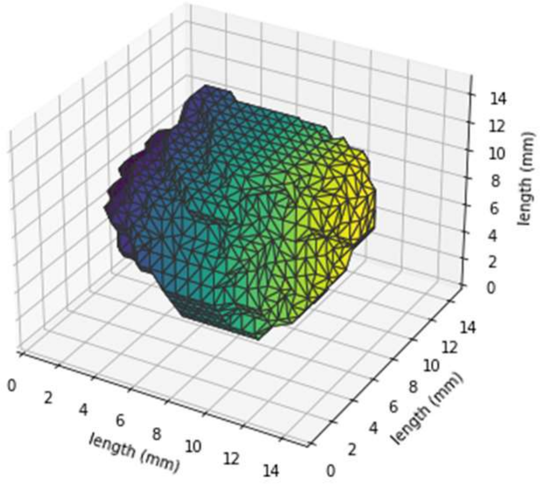
Classification

- Trained k-NN and FMD classifiers to classify between benign/malignant
- 75/25 split training and test
- Average over 50 iterations









Lung tumour classification using rank invariants

	Filtration	k-NN		MBD	
		Accuracy	AUC-ROC	Accuracy	AUC-ROC
Primary Benign vs Malignant	z-Rips	61.4	59.9	68.8	69.1
	x-Rips	64.4	64.2	67.9	68.2
	y-Rips	61.2	60.4	70.0	71.0
	Degree-Rips	63.3	61.8	**70.8**	**72.0**

	Filtration	k-NN		MBD	
		Accuracy	AUC-ROC	Accuracy	AUC-ROC
Primary Benign vs Malignant (with contrast)	z-Rips	**83.8**	**83.0**	76.9	76.8
	x-Rips	80.2	79.1	80.7	79.9
	y-Rips	79.6	78.5	80.0	79.3
	Degree-Rips	80.0	79.1	72.5	72.7

Previous studies using single-parameter persistent homology by Vandaele et al. (2023) achieved at most AUC-ROC of **67.7** for **data without contrast** and **78.0 with contrast**

Summary

Do rank functions and rank invariants perform well in inferential settings?

- We produced three different applications of **inferential and supervised learning tasks** to **real and simulated data**
- In all of them we observed an **improvement in performance** with respect to
  - Non-persistence-based methods
  - (If rank invariants were used) single-parameter based methods
- We believe that marrying FDA methods with rank invariants opens the door to many tools in TDA which are worth exploring in data-driven applications

Stability of rank functions

Matching distance

- Some stability results for the matching distance in size functions in the work of Landi and Frosini in the 1990s
- The matching distance can be extended to manifolds defined from vector fields (multiparameter persistence)
- Cerri et al. (2013) prove that the matching distance is stable under some conditions
- Landi (2018) shows that it is also stable with respect to the interleaving distance
- Kerber et al. (2019) show that for biparameter persistence, the matching distance is computable in polynomial time
- **It does not provide Hilbert structure**

L^p distance

$$d_{L^p}(f,g) := \left(\int_{\mathbb{R}^{2n+}} (f-g)^p d\omega\right)^{1/p}$$

- Provides Hilbert structure and is easily computable
- Known to have worse stability behaviour, but to what extent?
- Only study present in the literature by Skraba and Turner (2021) for *weighted metrics.*
- L^p distances are in fact rigorous metrics when there are no essential cycles.

Stability with respect to bottleneck distance

Definition

For β rank function and $\delta > 0$, the δ-*truncated rank function* is defined as

$$\beta_\delta := \beta \cdot \mathbb{1}_{\mathbb{R}^{2+}_\delta}$$

where $\mathbb{1}_{\mathbb{R}^{2+}_\delta}$ is the indicator function of the set $\mathbb{R}^{2+}_\delta := \{(x, y) \in \mathbb{R}^{2+} \ : \ y > x + \delta\}$.

Proposition (Bottleneck stability for truncated rank functions)

Let $1 \leq p < \infty$ and M be a p.f.d. persistence module with finite intervals in its barcode decomposition. For every $\delta > 0$, there exist $1 \geq \eta > 0$ and $K_{M,p} > 0$ such that any persistence module N satisfying

$$d_B(\mathrm{Dgm}\,(M), \mathrm{Dgm}\,(N)) < \eta$$

also satisfies

$$\left\| \beta_\delta^M - \beta_\delta^N \right\|_p \leq K_{M,p} \cdot d_B(\mathrm{Dgm}\,(M), \mathrm{Dgm}\,(N))^{1/p}. \tag{2}$$

In other words, the map $(\mathcal{D}, d_B) \to (\mathcal{I}_1, L^p)$ which sends each persistence diagram to its correspondent rank function is locally Hölder with exponent $1/p$.

Stability with respect to 1-Wasserstein distance

Theorem (Skraba and Turner (2021))

Let $f, g : K \to \mathbb{R}$ be monotone functions on a finite CW-complex K. Then

$$W_p(\mathrm{Dgm}\,(f),\, \mathrm{Dgm}\,(g)) \leq \|f - g\|_\infty.$$

Theorem (1-Wasserstein stability for rank functions)

Let $p = 1, 2$; and M be a p.f.d. persistence module with finite intervals in its barcode decomposition. Then there exists a constant $C_{M,p} > 0$ such that for any other p.f.d. persistence module N satisfying $W_1(\mathrm{Dgm}\,(M),\, \mathrm{Dgm}\,(N)) \leq 1$, we have

$$\left\| \beta^M - \beta^N \right\|_p \leq C_{M,p} \cdot W_1(\mathrm{Dgm}\,(M),\, \mathrm{Dgm}\,(N))^{1/p}. \tag{3}$$

In other words, the map $(\mathcal{D},\, W_1) \to (\mathcal{I}_1,\, L^1)$ which sends a persistence diagram to its corresponding rank function is locally Lipschitz, and the same map between the spaces $(\mathcal{D},\, W_1) \to (\mathcal{I}_1,\, L^2)$ is locally Hölder with exponent $1/2$.

Stability of rank invariants

Definition

Let $M, N : \mathbb{R}^n \to \mathrm{Vec}$ be modules decomposing in the intervals $\{J_j : j \in \mathcal{J}\}$ and $\{K_k : k \in \mathcal{K}\}$. The *p-Wasserstein distance* between M and N is defined as

$$d_{W_p}(M,N) = \inf_{\phi:\mathcal{I}\to\mathcal{K}} \left[\sum_{\phi(i)=k} d_I(\mathbb{I}^{J_i}, \mathbb{I}^{K_k})^p + \sum_{j\in\mathcal{J}\setminus\mathcal{I}} d_I(\mathbb{I}^{J_j}, 0)^p + \sum_{k\in\mathcal{K}\setminus\phi(\mathcal{I})} d_I(0, \mathbb{I}^{K_k})^p \right]^{1/p}$$

where ϕ ranges over all injections of subsets $\mathcal{I} \subset \mathcal{J}$ into $\mathcal{K}$.

Proposition

Let M and N be rectangle decomposable $\mathbb{R}^n$-persistence modules. Then, for $p = 1, 2$ there exist $c_{M,N,p,n} > 0$ such that

$$\left\| \beta^M - \beta^N \right\|_p \leq c_{M,N,p,n} \cdot d_{W_p}(M,N)^{1/p}.$$

Summary

What can we say about the stability of rank functions and rank invariants with L^p metrics?

- Rank functions with L^p metrics and barcodes with the bottleneck distance have very different stability behaviour, we need to exclude points close to the diagonal if we want to achieve a stability bound
- Leveraging new results concerning the stability of barcodes and the 1-Wasserstein distance, we provided stability bounds for rank functions with L^p metrics with respect to these
- We also were able to extend this results to multiparameter, rectangle indecomposable modules

Thanks for your attention!